Investigation of extended Y chromosome STR haplotypes in Sardinia

Investigation of extended Y chromosome STR haplotypes in Sardinia

Lacerenza D[a*], Aneli S[b,c], Di Gaetano C[b,c], Critelli R[c], Piazza A[b,c], Matullo G[b,c], Culigioni C[d], Robledo R[e], Robino C[a], Calò C[d]

[a] Dipartimento di Scienze della Sanità Pubblica e Pediatriche, Università di Torino, Italy

[b] Dipartimento di Scienze Mediche, Università di Torino, Italy

[c] HuGeF, Human Genetics Foundation, Torino, Italy

[d] Dipartimento di Scienze della Vita e dell'Ambiente, Università di Cagliari, Italy

[e] Dipartimento di Scienze Biomediche, Università di Cagliari, Italy

[*] **Corresponding author**: Dipartimento di Scienze della Sanità Pubblica e Pediatriche, Università di Torino, Corso Galileo Galilei 22, 10126 Turin, Italy

e-mail: lacerenza.d@gmail.com (D. Lacerenza)

Abstract

Y-chromosomal variation of selected single nucleotide polymorphisms (SNPs) and 32 short tandem repeat (STR) loci was evaluated in Sardinia in three open population groups (Northern Sardinia, n = 40; Central Sardinia, n = 56; Southern Sardinia, n = 91) and three isolates (Desulo, n = 34; Benetutti, n = 45, Carloforte, n = 42). The tested Y-STRs consisted of Yfiler® Plus markers and the seven rapidly mutating (RM) loci not included in the YFiler® Plus kit (DYF399S1, DYF403S1ab, DYF404S1, DYS526ab, DYS547, DYS612, and DYS626).

As expected, inclusion of additional Y-STR loci increased haplotype diversity (*h*), though complete differentiation of male lineages was impossible even by means of RM Y-STRs (*h* = 0.99997).

Analysis of molecular variance indicated that the three open populations were fairly homogeneous, whereas signs of genetic heterogeneity could be detected when the three isolates were also included in the analysis.

Multidimensional scaling analysis showed that, even for extended haplotypes including RM Y-STR markers, Sardinians were clearly differentiated from populations of the Italian peninsula and Sicily. The only exception was represented by the Carloforte sample that, in accordance with its peculiar population history, clustered with Northern/Central Italian populations.

The introduction of extended forensic Y-STR panels, including highly variable RM Y-STR markers, is expected to reduce the impact of population structure on haplotype frequency estimations. However, our results show that the availability of geographically detailed reference databases is still important for the assessment of the evidential value of a Y-haplotype match.

Keywords

Y chromosome, Y-SNP, Y-STR, Sardinia, Italy

1. Population

Sardinia is a well-known genetic isolate within Europe, with both archeological/historical data and modern/ancient DNA analysis suggesting population continuity since the Neolithic [1]. Several factors, including genetic drift, founder effects and positive selective pressure, driven by endemic malaria, are major contributors in determining the distinct outlier position of Sardinians. However, it is still debated whether genetic drift and founder effects also acted within the island causing genetic heterogeneity among sub-regions [2-5].

The possible presence of population structure must be taken into account in forensic Y-chromosomal short tandem repeat (STR) analysis, in order to adopt adequately detailed haplotype reference databases for match calculations. The need for population substructure corrections may be partially overcome by applying rapidly mutating (RM) Y-STRs [6], which have been recently integrated in commercial kits (e.g. DYS570, DYS576, DYS627, DYS518, DYS449, and DYF387S1 in the YFiler® Plus kit) with the purpose to achieve nearly individualization of male Y-STR profiles. Due

to their high mutation rates, RM YSTRs were shown to minimize signals of population structure, at least at macrogeographic level [7].

Because of their peculiar and, to some extent, unclear genetic features (in terms of homogeneity/heterogeneity), Sardinians can be considered an ideal population to test the impact of RM Y-STRs on haplotype match calculations. Hence, the aim of the present study was to investigate a Sardinian population sample, which had been previously partly typed for the standard AmpF*l*STR® YFiler® loci [8], by means of an extended set of Y-STRs encompassing the YFiler® Plus loci and the seven RM Y-STRs not included in the YFiler® Plus kit.

2. Samples

Samples from 308 Sardinian males were collected by means of venipuncture or buccal swab. Donors belonged to three open populations from Northern (n = 40), Central (n = 56) and Southern (n = 91) Sardinia, and the three isolated populations of Benetutti (n = 45), Desulo (n = 34) and Carloforte (n = 42) (Supplementary material, Figure S1). Desulo [9] and Benetutti [10] are geographic isolates, given their secluded position in mountain areas of inner Sardinia (Barbagia and Goceano, respectively). Carloforte was chosen as an outgroup, since its population takes origin from migrants of Northern Italian origin, who were relocated from the Genoese settlement of Tabarka (Tunisia) in 1738, and still speak a distinctive archaic form of the Ligurian dialect [11].

All individuals were unrelated, apparently healthy, born and resident in the selected villages, or areas, for at least three generations. The study was reviewed and approved by the University of Cagliari Ethical Committee and all voluntary participants read and signed an informed consent form.

3. Extraction, PCR amplification, genotyping and statistical analyses

Genomic DNA was extracted using the QIAamp DNA Mini kit (Qiagen).

Single nucleotide polymorphism (SNP) genotyping was performed by minisequencing of 18 markers defining the major haplogroups and subhaplogroups on the Y-chromosome human phylogeny in

Europe. Y-SNPs were assembled in two multiplexes (including M35, M89, M170, M172, M9, M45, M173 and M91, M181, M216, M174, M96, M201, M52, M214, respectively), as described by Onofri et al. [12]. Typing of marker M26 in M170 positive individuals was done by restriction enzyme analysis. Y-haplogroup nomenclature followed the reference tree available online at http://www.phylotree.org/Y [13].

The following Y-STR loci were amplified with the Yfiler® Plus amplification kit (Thermo Fisher Scientific): DYS576, DYS389I/II, DYS635, DYS627, DYS460, DYS458, DYS19, YGATAH4, DYS448, DYS391, DYS456, DYS390, DYS438, DYS392, DYS518, DYS570, DYS437, DYS385, DYS449, DYS393, DYS439, DYS481, DYF387S1, DYS533. The remaining RM Y-STR loci (DYF399S1, DYF403S1ab, DYF404S1, DYS526ab, DYS547, DYS612, and DYS626) were amplified in a single multiplex PCR assay: PCR primers and amplification conditions were those described by Robino et al. [14], the only modification being a change of dye label for locus DYS626 (from 6-FAM to TAMRA) in order to avoid overlapping between marker-specific fluorescent signals. Detection and separation of PCR products were carried out using the ABI Prism 3500 Genetic Analyzer (Thermo Fisher Scientific) and Osiris software version 2.7 [15]

Statistical calculation of standard diversity indices, pairwise genetic distances (F_{ST}) and analysis of molecular variance (AMOVA) was performed with ARLEQUIN software version 3.5 [16]. Multidimensional scaling (MDS) analysis of population matrixes of Slatkin's linearized genetic distances was performed using the AMOVA tool from the YHRD website [17] and R package *stats* v.3.3.0 [18].

4. Quality control

This manuscript follows the guidelines for the publication of population data indicated by the journal [19]. Accordingly, Y-SNP and Y-STR data for loci included in the Yfiler® Plus kit were preliminary submitted to the YHRD (http://www.yhrd.org) and assigned the following accession number: YA003993.

For supplementary RM Y-STR markers, quality control measures as described in [14] were adopted.

5. Results

Y-chromosomal haplogroups and haplotypes observed in the 308 samples from Sardinia are listed in Supplementary material, Table S1.

6. Other remarks

The present study represents a continuation of an earlier investigation of Y chromosomal variation in Sardinia [8]. Besides the addition of Y-SNPs and new Y-STRs recently introduced in forensics, as a further improvement to [8] the Southern Sardinia sample was expanded to include previously untested areas.

Distribution of Y-SNP haplogroups (Supplementary material, Table S2) was in line with previous studies of the Sardinian population [20,21]. In particular, haplogroup I2-M26 was the most represented in the sample, with a frequency of 34.7%, compared to 0.6-1.4% observed in other Italian populations [22]. Frequencies of haplogroup I2-M26 within Sardinian populations ranged between 27.5% (Northern Sardinia) and 48.9% (Benetutti). Only in Carloforte I2-M26 was rare, being observed in a single individual (2.4%), and haplogroup distribution was similar to that in continental Italy, with R1 by far the most common haplogroup (54.8%) [22].

Considering the whole set of tested markers, including both YFiler® Plus loci and supplementary RM Y-STRs (referred to, from now on, as "extended haplotype"), a total of 307 different haplotypes were observed. Haplotype diversity (*h*) values are summarized in Supplementary material, Table S3. Addition of new Y-STRs increased power of discrimination in Sardinia (*h* = 0.99997), compared to the standard 17 AmpF*l*STR® YFiler® loci (*h* = 0.998) [8], but did not provide complete male individualization, with a single extended haplotype found in two subjects from different sampling sites in Central Sardinia. This was also the only 13 RM Y-STR haplotype match found in our Sardinian sample. When limiting the analysis to YFiler® Plus loci, 302 unique haplotypes were seen,

with six haplotypes found twice, always within single populations (*h* = 0.99987). For YFiler® Plus markers the lowest diversity levels were observed in Benetutti (*h* = 0.99803) and Desulo (*h* = 0.99824).

Results of AMOVA conducted on extended haplotypes are summarized in Supplementary material, Tables S4. When considering the isolates of Benetutti, Desulo and Carloforte as separate groups, AMOVA suggested the presence of genetic heterogeneity within the island (F_{ST} = 0.030; $p < 0.00001$; 2.05% of the observed variation among population groups). The percentage of variation among groups was reduced (0.40%), but still significant (F_{ST} = 0.022; $p < 0.00001$), after the outgroup of Carloforte was excluded from calculations and the two remaining isolates were grouped according to their geographic position (Benetutti with Northern Sardinia, Desulo with Central Sardinia). Following the exclusion from the analysis of the Benetutti and Desulo samples, variation among the three open population groups became negligible (0.00%), with F_{ST} = 0.008 ($p > 0.5$).

Population pairwise F_{ST} distances always reached statistical significance for the isolates of Desulo and Carloforte. On the contrary, the only significant comparison between Benetutti and an open population was with Southern Sardinia (Supplementary material, Table S5).

MDS analysis within Sardinia is displayed in Supplementary material, Figure S2. A clear separation between the cluster of Northern, Central and Southern open populations, and isolates (Desulo and Carloforte, in particular) could be observed.

Comparison with other European populations genotyped for YFiler® Plus loci available in the YHRD database (Release 51) confirmed Sardinians as outliers within the European genetic landscape (Supplementary material, Figure S3).

No matches were seen between Sardinian YFiler® Plus haplotypes observed in the present study and previous YFiler® Plus studies of Italian populations, consisting of 150 individuals from Northern Italy [23], and 203 individuals from the whole Italian peninsula including Sardinians (n = 51) [24]. Similarly, no RM Y-STR match was found after comparison with a large database (n = 1509) of individuals from the Italian peninsula and Sicily [14]. A clear differentiation between Sardinians and

Italians was evident from MDS plots obtained with both sets of markers (Supplementary material, Figure S4). The only exception was represented by the Carloforte sample that, in accordance with its peculiar population history [11], clustered with Northern/Central Italian populations.

The obtained results confirmed that even the inclusion of RM Y-STRs did not completely erase the outlier position of Sardinia within the Italian genetic landscape, highlighting the necessity to integrate with Sardinian-specific data the Italian Y-haplotype reference databases currently available for forensic match calculations. Distribution of extended Y-STR haplotypes was quite homogeneous between pooled samples belonging to different Sardinian sub-regions (North, Center, South). However, signs of sub-structuring previously observed for standard YFiler® loci [8] were confirmed in communities where a long history of isolation, either for geographical (Desulo) or historical (Carloforte) reasons, is documented [9,11]. Therefore, it is important that Sardinian Y-haplotypes complementing Italian reference databases include microgeographically detailed data, in order to detect finer scale patterns of variation which may be relevant for the interpretation of Y-STR evidence.

References

[1] M. Sikora, M.L. Carpenter, A. Moreno-Estrada, B.M. Henn, P.A. Underhill, et al. Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe, PLoS Genet. 10 (2014) e1004353.

[2] C. Di Gaetano. G. Fiorito, M.F. Ortu, F. Rosa, Guarrera S, et al., Sardinians genetic background explained by runs of homozygosity and genomic regions under positive selection, PLoS One 9 (2014) e91237.

[3] L. M. Pardo, G. Piras, R. Asproni, K.J. van der Gaag, A. Gabbas A, et al., Dissecting the genetic make-up of North-East Sardinia using a large set of haploid and autosomal markers, Eur. J. Hum. Genet. 20 (2012) 956-964.

[4] I.S. Piras IS, A. De Montis, C.M. Calò, M. Marini, M. Atzori, et al., Genome-wide scan with nearly 700,000 SNPs in two Sardinian sub-populations suggests some regions as candidate targets for positive selection. Eur. J. Hum. Genet. 20 (2012)1155-1161.

[5] G. Pistis, I. Piras, N. Pirastu, I. Persico, A. Sassu, et al., High differentiation among eight villages in a secluded area of Sardinia revealed by genome-wide high density SNPs analysis, PLoS One 4 (2009) e4654

[6] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, et al., A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, Forensic Sci. Int. Genet. 6 (2012) 208-218.

[7] K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, et al., Towards male individualization with rapidly mutating Y-chromosomal STRs, Hum. Mutat. 35 (2014) 1021-1032.

[8] R. Robledo, A. Mameli, C.M. Scudiero, G. Vona, L. Corrias, et al., Non-random distribution of 17 Y-chromosome STR loci in different areas of Sardinia, Forensic Sci. Int. 16 (2015) 26-28.

[9] S. Sartoris, O. Varetto, N. Migone, N. Cappello, A. Piazza, et al., Mitochondrial DNA polymorphism in four Sardinian villages, Ann. Hum. Genet. 4 (1988) 327-340.

[10] M. Capocasa, P. Anagnostou, V. Bachis, C. Battaggia, S. Bertoncini, et al., Linguistic, geographic and genetic isolation: a collaborative study of Italian populations, J Anthropol Sci 92 (2014) 1-32

[11] R. Robledo, L. Corrias, V. Bachis, N. Puddu, A. Mameli, et al, Analysis of a genetic isolate: the case of Carloforte (Italy), Hum. Biol. 84 (2012) 735-54.

[12] V. Onofri, F. Alessandrini, C. Turchi, M. Pesaresi, L. Buscemi, A. Tagliabracci, Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs, Forensic Sci. Int. 157 (2006) 23–35.

[13] M. van Oven M, A. Van Geystelen, M. Kayser, R. Decorte, M.H. Larmuseau. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome, Hum. Mutat. 35 (2014) 187-191.

[14] C. Robino, Ralf A., S. Pasino, M.R. De Marchi, K.N. Ballantyne, et al., Development of an Italian RM Y-STR haplotype database: Results of the 2013 GEFI collaborative exercise, Forensic Sci. Int. Genet. 15 (2015) 56-63.

[15] R.M. Goor, L. Forman Neall, D. Hoffman, S.T. Sherry, A mathematical approach to the analysis of multiplex DNA profiles, Bull. Math. Biol. 73 (2011) 1909-1931.

[16] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564-567.

[17] S. Willuweit, L. Roewer L., The new Y Chromosome Haplotype Reference Database, Forensic Sci. Int. Genet. 15 (2015) 43-48

[18] R_Core_Team (2015). "R: A Language and Environment for Statitical Computing"

[19] A. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, et al., New guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 7 (2013) 217-220.

[20] P. Francalacci, L. Morelli, A. Angius, R. Berutti, F. Reinier, et al., Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny, Science 341 (2013) 565-569.

[21] P. Francalacci, D. Sanna, A. Useli, R. Berutti, M. Barbato, et al., Detection of phylogenetically informative polymorphisms in the entire euchromatic portion of human Y chromosome from a Sardinian sample, BMC Res. Notes 8 (2015) 174.

[22] A. Boattini, B. Martinez-Cruz, S. Sarno, C. Harmant, A. Useli, et al., Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata, PLoS One 8 (2013) e65441.

[23] E. Ottaviani, S. Vernarecci, P. Asili, A. Agostino, P. Montagna, Preliminary assessment of the prototype Yfiler® Plus kit in a population study of Northern Italian males, Int. J. Legal Med. 129 (2015) 729-730.

[24] C. Rapone, E. D'Atanasio, A. Agostino, M. Mariano, M.T. Papaluca, et al., Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler(®) Plus kit) in an Italian population sample, Forensic Sci. Int. Genet. 21 (2016) e1-5.